Data Trust for the Royal Borough of Greenwich and Greater London Authority

Pilot 1 -
Sharing Cities
Transport and Energy Case Studies
Introduction

The Greater London Authority ("GLA") and the Royal Borough of Greenwich ("RBG") have jointly expressed an interest in trialling the concept of a data trust as a solution to the concerns held by themselves and data providers and users around the sharing of data, and general problems they experience with data sharing. As a general point of principle, both are interested in using the sharing of data to improve efficiency within the individual boroughs and in London as a whole.

In order to act as a proof of concept for them, and as areas which they think they might benefit from a data trust, RBG have provided access to two pilot project examples, and their stakeholders, to trial the notion of a data trust.

RBG, through the European Union funded Sharing Cities program, is the borough within which the two example use cases are being trialled. The role of the GLA is in providing data analysis and storage capacity and to potentially liaise with, and promote implementation of, a data trust model across all 33 boroughs if it is found to be a successful model for RBG. The ultimate aim with such a London-wide platform is to enable information to be shared across boroughs to improve decision making and prevent the same mistakes being made from one borough to the next. The GLA also acts as an intermediary and operator of the London Data Store ("LDS"), who will provide some of the data analytics and potentially the platform for storing some of the data. The Future Cities Catapult ("FCC") has the purpose of engaging with citizens and building a service layer on the back of data collected by RBG. The interviewee spoken to sees a data trust as a tool that has the potential to help cities operate in a smarter way for the benefit of its citizens1.

Data sharing arrangements already exist, within the wider legal landscape, in the form of Swift2 and Bolero3, both of which allow the sharing of data across disparate parties for a common purpose. The aim will be for the GLA and RBG to achieve something similar except within a local authority and London-wide setting. Additionally, this Pilot seeks to take data sharing further than the two aforementioned examples as RBG and the GLA wish to able to share large datasets, likely in real-time, and in a variety of different formats. Swift and Bolero are limited, in that they are primarily data messaging applications that contain data within that message. This means that they are not sharing large data sets and, once the message has been sent, that is the end of the data sharing arrangement.

1 https://futurecities.catapult.org.uk/2016/01/20/the-future-cities-catapult-to-lead-on-citizen-engagement-for-sharing-cities-lighthouse-programme/
2 https://www.swift.com/
3 http://www.bolero.net/home/carriers-and-logistics/
Transport Use Cases

The first of the two use cases revolves around sharing of transport data in order to encourage more journeys to be taken by electric vehicles. This goal is echoed across RBG, the GLA, and by the Mayor of London who is trying to reduce the number of trips taken by cars in London and increase the number of trips taken by low carbon transport methods.

To this end there has been an investment in a number of companies to help achieve this goal. There is an electric car club being provided by Enterprise Holdings, electric vehicle charging points being provided by BluePointLondon Limited trading as Source London and smart parking sensors provided by Smart Parking Ltd. Each of these individually provide a service to RBG but the aim of the data trust is to allow cross sharing of data that can yield even greater gains. Examples of this might be the smart parking sensors’ being able to determine if a car is parked in a space, cross-referenced against whether an electric vehicle is parked there or in an charging station. This means that RBG could be aware if a regular car is being parked in an electric vehicle car club designated space. In turn, this would mean that they could optimise the use of such spaces by electric vehicles, influence officers patrolling routes in order to take enforcement action against the users of regular cars misusing such spaces, which in the long term should reduce the number of individuals who are misusing these services and thereby hopefully increase the parking space uptake by electric vehicle car club users.

This is a goal that is of great interest to the GLA. However, RBG will be keen to use such data on whether there is misuse, to inform parking enforcement officers and increase fines with the ultimate aim of increasing compliance with parking rules. Currently the “technology” used to map data for all traffic wardens is paper based. Until recently as well, the way of determining where a car had been incorrectly parked would be described textually rather than by use of a map.

Energy Use Case

The second of the two use cases relates to heating and energy consumption and how energy usage can be optimised in council owned blocks of flats. The approach is twofold: first, there are smart heating sensors being installed in a communal heating system that measure their energy usage; secondly, temperature and humidity sensors are based in the flat itself that can be used to anticipate future heat demand. The aim of the sensors is to gather data and relay this back to the GLA and London Data Store (LDS), where the Sustainable Energy Management System, built by Imperial College London and Siemens can access the data for analysis and results are then sent back to the communal heating system to make recommendations on parameters for the heating system which can be changed accordingly. These recommendations can be changed based on energy use throughout London, outside temperatures and other factors.

Analysis

Given the various organisations involved, there are differing viewpoints on what a data trust should look like and what it should aim to achieve. No party has a firm idea of the form that a data trust should take. In part this is due to there currently being no defined legal standard to compare it with, as the idea of a data trust is a new concept, and also because each stakeholder has various concerns around data and they see a data trust as having the potential to be a catch-all solution to solving these problems.

https://ecarclub.co.uk/
https://www.sourcelondon.net/
https://www.smartparking.com/keep-up-to-date/case-studies/city-of-westminster-london
Common Themes

There are however some similarities across all of the interviewees with whom we have had the opportunity to speak. All see a data trust as being a way to enable effective and, more importantly, responsible sharing of data across various organisations for the benefit of those organisations and the wider public as a whole. All interviewees see that the data trust format should not be limited to the benefit of just the contributors that provide the data but also to help the wider community.

Despite being public, or a publicly-funded body in the case of the FCC, all see the need for there to be some form of business model in order to generate revenue to sustain the data trust and also to provide some recompense to the data providers and organisations including the GLA, RBG and any other private body involved in providing its data, be that an organisation or even potentially an individual. All those interviewed however were equally against any model that revolved around the wholesale selling of data or targeted advertising, this opposition reflects general public opinion.

All interviewees think there should be clear consent and engagement from data providers, with any consent form being simple to understand, and making clear to providers exactly what their data will be used for. All those involved are of the opinion as well that there should be a single general consent that will cover all permitted uses of the data, and not to have multiple consents for each and every possible permutation of use, in order to prevent consent fatigue and administrative complexity.

All interviewees also echo similar concerns about the current method used for data sharing being inadequate. A predominantly used system of bilateral data sharing agreements between each pair of data sharing organisations is a time consuming and involved process that is generally inflexible. It is difficult to know what other boroughs or organisations are doing and this can lead to duplication of effort and difficulties in innovators approaching London boroughs if they are looking to implement a product wider than just one individual borough.

Interestingly, there is a universal consensus that there should be careful restrictions as to who can access any data shared with the data trust. All parties see that data should only be accessed for the reason for which permission was granted by the data provider, a stipulation which is mirrored under the provisions of GDPR, and that there should be strict controls preventing misuse of the data. There is also a fear amongst the interviewees of the data trust not being any more of an effective system of sharing data than the current data sharing agreement method, and a fear that data could be accessed by police, immigration or security services either through a warrant or court order or else through a simple informal request from those services, without proper consideration over whether the request should be acceded to. Whilst such bodies could request such data from the data provider, in situations like with the energy use case, where the sensors could be used to determine when the heating is on, and therefore when someone was in the property, this could influence decisions about the best time of day to arrest data providers.

---

8 https://www.marketingweek.com/2017/06/29/arrogance-brand-purpose-distrust-ads

9 Article 4(11) of GDPR
Suggested Forms

Whilst there is a broad consensus over the general principles, risks and practical concessions (such as acceptance of the possibility of the trust having to hold personal data and therefore comply with GDPR) a data trust might make, there is an understandably diverging view as to what the legal format of a data trust should be. Most seem to agree that centralising data storage in a London wide data trust would be undesirable due to the unwieldy nature of having such a large amount of data stored in one place. Additionally, if personal data is being stored centrally this places a large responsibility on the part of the data trust to comply with GDPR on behalf of numerous disparate individuals who would have their data stored within the data trust.

Various possible alternatives are possible. First, to have individual data trusts for each borough with a central register to indicate what data is held by each borough and which may or may not hold meta-data, analysed data or other pattern data. Secondly, albeit a form which does not conform to current general thinking of a data trust, is to have the data trust be an adviser, either advising on the ethical sharing of data, or assisting with data sharing contractual documents. With this adviser model, the data trust could also act as a go-between for interested parties to approach one another and then to help facilitate the sharing of data.

Below are headings that identify the different aspects that need to be considered when constructing a data trust. These mirror the headings within the general legal report and address each legal point that must be considered when constructing a data trust. The analysis will also cover which suggested element from each of the headings would likely be the most appropriate for the individual stakeholders and the two use cases as a whole.

"there is an understandably diverging view as to what the legal format of a data trust should be"
Possible Legal Structures

Some approaches

In terms of possible legal structure, it would seem that the contractual model would represent no improvement on the system that the GLA and RBG currently use. The difficulty that the GLA and RBG currently face using multiple data sharing agreements are that they are time consuming to negotiate, inflexible and limited in scope. It would not be ideal to replicate such an approach more widely to meet the GLA and RBG's data sharing needs. The typical potential forms would be a standard form contract with all agreements mirroring the provisions in one another, or a single contract to which all data providers and data users could sign up. These would likely limit access to data more so than, say, an organisational data trust model which both ensures compliance with the agreement(s) or which can grant access to data. This would again push the responsibility of entering into such data sharing agreements onto the GLA and RBG and it would be more difficult for parties interested in data to access than if an application was made to a central body dedicated to sharing data in a responsible manner.

On the surface, the traditional equitable trust model might seem attractive, as it meets the interviewee requirements of a standard set of terms (the legal rules) that would apply to all those creating a data trust and that has inbuilt mechanisms to ensure compliance and a governance structure. However, the fact that potentially time-consuming or unviable law reform would need to take place before this method could be used, would make it wholly unsuitable for RBG and the GLA who are developing these use cases in real-time and who need a model that can be implemented as soon as possible in order to reap immediate gains. It should also be noted that whilst it is the GLA and RBG as public bodies driving forward the implementation of a data trust, several interviewees have emphasised the need to promote London's economic growth and for it to become a leader in data sharing and AI development. This could not be achieved if the government had to take years to make any reforms before an equitable data trust model could be utilised.

The suggested public model of a data trust, wherein a hypothetical public regulator ensures compliance with data trust rules, suffers from a similar issue. If this was going to happen it would need to come from central government as, whilst the GLA and RBG are keen to reap the benefits of having a data trust sharing model, they have generally indicated that they do not have an appetite to take on the role of regulator for London data trusts. Again, the issue is central government having the political wherewithal and motivation to put this in place. Even if motivated to do so, it would likely take longer to implement than the ideal delivery date for RBG and the GLA who are looking for an immediate solution in order to enable them to capitalise on the sharing of data.
Corporate/organisational models

Therefore, the corporate or organisational model for a data trust is likely hold the most appeal to the GLA and RBG. Despite a clear preference in the general legal report for the Community Interest Corporation ("CIC"), this or the Limited Company ("LC") models are both viable options. The great appeal of the corporate model for the GLA and RBG would be that corporate or organisational structures are already commonplace and there are established ways by which these can be governed through a board of directors.

The partnership model, likely in the form of an LLP due to its limitation of liability, and the company limited by guarantee are both potential options as well. However, because each partner or member of these organisations gets an equal voting share in the organisation, this might make them less viable, as organisations that contribute a large amount of data are likely to want a greater say in the running of the data trust than those who only contribute a small amount of data.

Considering the needs of RBG, the GLA and individual providers, there are two viable models that could be used depending on the approach preferred. Both are forms of corporate structure, one being as a LC and one a CIC. The advantage of favouring a corporate structure is that it has its own separate legal identity capable of holding assets and having legal rights assigned to it. So, for example, if each data provider chooses to license its data to the data trust (see below), then the licence can be to the corporate body that makes up the data trust, rather than to any one individual prospective data user.

This means that it meets the stakeholder requirement to have the data trust take the form of a third party organisation, thereby ensuring that data is not licensed to one party over another which could skew the balance of power. This would be because the independent third-party organisation is unbiased and therefore will not offer preferential licence terms to one prospective data user over another or arbitrarily refuse access, although it is likely that the data provider will still wish to set parameters over the types of data users who can access its data. This, additionally, centralises the data and maintains an organisational structure that can not only facilitate between parties the sharing of data, but also act as an overseeing body, which has the potential to be independent depending on the governance structure of the trust.

Both forms of company structure also share the quality of having pre-established structures, using an appointed board of directors and shares in the company to balance the needs of the individuals. Differing amounts of shares in the company structure could be granted to reflect differing levels of contributions to the data trust. A corporate structure is also a very simple and established process to set up, and flexible in its form. Additionally, the governance documents of the corporate entity can be drafted so that the directors, who will be responsible for the day-to-day operation of the trust, will be required to act in a way that supports the responsible sharing of data. This means that they are effectively acting as the "trustees" of the data trust, to use the equitable trust law comparison, in their role as directors.

The differences between the LC and CIC are important to note and will influence which model suits the stakeholders best.
Limited Company (LC) Model

A LC is, in most cases, set-up for the purpose of making a profit and not for wider social good. It is possible to have a LC, whose main goal is to provide benefit for the public good or for a certain set of beneficiaries of the trust, by amending its governance documents, being its Memorandum and Articles of Association, to specifically state the positive purpose that the LC needs to pursue.

These would ensure that within the company’s constitution there is an in-built company purpose that states the company is only to act in a way that benefits the public good through the ethical sharing of data, that a certain percentage of revenue is allocated to a public good or in the event of winding-up of the company that the company’s resources are used for some public good. In the case of a data trust, such good could be supporting other data trusts or organisations that support an ethical data sharing arrangement.

Each data provider could be a shareholder in the company, with a greater amount of shares being given to those contributing the most valuable data or resources. Any profits made by the data trust (eg. from fees for access to data), therefore, would be distributed accordingly between all parties so that those that contribute the greatest value of data receive the most benefit as holders of the most shares. It could even be possible to have different classes of share with different rights depending on various stakeholder contributions or potentially levels of access. RBG and the GLA as both organisation founders and contributors of resources, in the form of data and the resources of the LDS, Imperial College London and Siemens (the latter two being organisations involved in the analysis of gathered data), could all be issued shares and therefore benefit from the founding of the data trust.

[10] Article 4(11) of GDPR
Valuing data contributions

The first of these is valuing data contributions. One data provider might contribute a large volume of data that entitles them to a lot of shares, but this may be data that data users are not interested in accessing. This would therefore unfairly prejudice those that contribute useful data of a smaller amount, as the dividend would be skewed towards the data provider who is providing large amounts of data but who in practical terms is contributing less. This would lead to feelings of unfairness on the part of a commercial actor providing its data who would want to see a return on its contribution which, in turn, could make them less likely to want to contribute their data in the first place. Trying to mitigate these by attempting to pre-determine the usefulness of the data would be difficult as well as it is very challenging to predict market trends and tastes to accurately put a price (i.e. how many shares to issue) on the value of the data before it is licensed to the data trust.

This could, however, be controlled by using the following approach:

- part of profits distributed as dividend based on volume of data contributed (or quantity used by data users, as some contributed data might not be used); and

- the remainder distributed by the board in proportion to value to data users – a methodology to assess this would have to be devised, but it could even be as simple as asking users to rate usefulness, or as complex as engaging consultants to value usefulness.

Additionally, a LC is still typically associated with profit maximization and therefore does not lend itself naturally to the kind of prosocial activities that the GLA and RBG are typically seeking to promote. It can be included in a company’s articles that they should act in a certain way (e.g. some prosocial provisions). However, entrenchment of these is not possible, because a company can amend its articles through a 75% majority of the shareholders.

This means that the purpose of a data trust could change drastically with the consent of some (but not all) of the shareholders. Such a move, whilst unlikely, has the potential to give the public a fearful impression as to the motives of a data trust, something that was of particular concern to RBG and the GLA.

Also, if the data trust was large, with many data providers, having so many shareholders with new ones being added every time a new data provider joins the data trust, could again be administratively difficult. Further, if a shareholder pulls their data from the trust, there might be the added administrative difficulty of having to reassess the Board representation to ensure it accurately reflects representatives from the remaining data providers. Additionally, if shares are being bought back from a data provider when they remove their data from the trust, then the company will be liable for stamp duty (although this could be a nominal amount to reflect that the buyback of shares does not represent a loss to the data provider as they would remove their data from the trust simultaneously to this).
Value returned under License

The alternative to having each shareholder being issued shares for the data they are providing, is to have a nominal group of individuals or organisations as shareholders, but the majority who provide their data to the data trust will be recompensed for providing their data through a payment made under the license agreement through which the data provider provides its data to the trust, entered into between the data provider and the data trust. In this model, the board would still manage the everyday operation of the data trust. Each data provider would licence their data to the trust and would receive some proportion of the money paid, either through a general subscription (similar to how Netflix or Spotify operate) or on a per access basis, when a data user seeks data from the particular data provider. This means that there is a fair reward for data providers based on the demand for their data. Whilst RBG and the GLA are of an opinion that it seems likely that there would need to be some charging for data in order to provide revenue for the continued management and running of the data trust, and so that some financial return can be made to the data providers licensing their data to the data trust, they would like to primarily see prosocial benefits that will meet the Mayor of London’s overarching goal to reduce carbon emissions in the city and increase the number of journeys being taken by public transport and electric vehicles. This second model could be strengthened through provisions around the composition of the Board, with seats reserved for representatives of any relevant social interests.

As mentioned in the general legal report however, a director could be in breach of his/her director’s duties if they do not act in such a way that promotes the success of the company. One example of this might be the potentially proposed discount that could be given to data user organisations that are charities or have a prosocial purpose, thereby not receiving full market value for access to the data. The test of what is promoting the companies’ best interests is subjective\(^5\) and a director needs to show their rationale in making a decision and that it was “bona fide in the interests of the company.”\(^6\) For this pilot, the directors of the LC data trust could be in breach of their duties as directors, unless they were promoting the success of the company for the benefit of its shareholders, employees and other stakeholders in the company\(^7\) (thereby potentially precluding prosocial benefits of promoting general data sharing). One way of remedying this is to have a specific prosocial purpose stipulated in the company’s articles but these will need to be amended if the scope of the data trust changes, as otherwise there will likely be some ambiguity as what is permitted by the articles. Additionally, a data provider that is a corporate body itself should also be careful about breaching its director’s duties in licensing data to the trust, unless it is receiving a benefit by doing so or can argue that such an action falls under any corporate social responsibility policy\(^8\) or some other justification, such as where the loss in revenue might be slight, but there is a large reputational or advertising gain.

\(^{15}\) Hellard and another (Liquidators of HLC Environmental Projects Ltd) v Carvalho [2013] EWHC 2876
\(^{16}\) Re Southern Counties Fresh Foods Ltd [2008] EWHC 2810
\(^{17}\) Dawson International Plc v Coats Paton Plc [1989] BCLC
One potential justification is in the licensing model (see below), where data is being licensed to the trust that might not otherwise be used. In the transport use case for example, the car club might not be getting any pecuniary benefit from their analytics data, however by contributing it to the data trust, there is a potential financial return on this to the company from data that would otherwise not be revenue generating. As referred to above, sharing data and promoting the ethical sharing of data is something that could be included under an organisation’s corporate social responsibility obligations. This is covered in various UK wide legal and policy initiatives to encourage companies’ do this. For premium listed companies, the UK Corporate Governance Code 2018 encourages companies to contribute to the wider society. For institutional investors there is the UK Stewardship Code which encourages investors to consider how they will engage with investee companies over things such as environmental risk. In terms of EU initiatives there is a directive that requires large EU listed companies of more than 500 employees to draw up an annual statement relating to environmental, social and employee matters.

The key point to note is that such attempts to encourage corporate social responsibility seek to promote specific goals. As a data trust is a new concept it would likely be the responsibility of the courts to decide whether the duty in the UK Corporate Governance Code 2018 of a company to contribute to wider society, would give some leeway to directors to not promote the success of the company (in financial terms). It should also be noted that the UK Corporate Governance Code 2018 applies to premium listed companies and would not apply to other organisations such as a limited company.

---

19 https://www.frc.org.uk/getattachment/88bd8c45-50ea-4841-95b0-d2f4f48069a2/2018-UK-Corporate-Governance-Code-FINAL.PDF

20 https://www.frc.org.uk/getattachment/dff25bf9-998e-44f6-a699-a697d932da60;=aspx

CIC Model

The workaround to enable a focus on prosocial activity whilst also providing some return on investment to the data providers is using the CIC model discussed in the general legal report. If the data trust does not use a model where distributions are made through share dividends, participating data providers could still receive a fee for their data being accessed under the agreement that licenses the data to the CIC (meaning that the 35% cap on dividends for distributable profits is not an issue).

The key benefit of using the CIC model from the point of view of GLA and RBG is that it has a demonstrable community-minded and prosocial purpose that needs to be complied with otherwise the CIC regulator, the Office of the Regulator of Community Interest Companies, can enforce compliance with its stated prosocial aim.

One of the key points in the mind of the GLA and RBG is that the model of the data trust is something that should persuade data providers, and the wider public, to have faith that their data will not be exploited. Such a sentiment was referred to in multiple interviews with both representatives from the GLA and RBG. If the form of the data trust is therefore not an inherently profit seeking one, unlike a standard, unamended LC, which the public generally distrust, this could go some way to achieving this goal. One of the repeated fears of the GLA and RBG is abuse by the data trust by the mass commercialisation of data, such as for targeted advertising.

A CIC data trust could be particularly relevant for the energy use case, where the data trust could potentially be handling tenants’ sensitive personal data, as a community minded data trust legal structure would likely encourage more confidence than the LC model. In the form the transport use case is currently in, it would potentially benefit most from the LC model, as there are three commercial partners in addition to the GLA and RBG. Because the commercial organisations’ role is revenue generation, a LC model might be better suited. There is no need to justify a prosocial purpose which might require used of a CIC. If though, as the GLA and RBG suggest, the data from the transport use case is shared more widely or more providers join the data trust, for example “Transport For London” (“TFL”), then the CIC model could be more appropriate. The main determining factor as to whether the CIC or LC model is appropriate is whether the data sharing benefits are limited to commercialisation, and therefore an LC model is best, or whether there is to be an overarching prosocial purpose behind the data trust, in which case the CIC model would be more appropriate. There will likely have to be some sort of mechanism for either payment for access to data or receiving some other benefit in kind.

---

22 Community Interest Company (Amendment) Regulations 2014

23 Sections 45-48 Companies Audit, Investigations and Community Enterprise 2004

The GLA, and to a lesser extent RBG, have suggested that rather than have one London-wide data store, there could be several smaller ones with the GLA, through the LDS, acting as a central register to inform prospective data users’ what data is being stored where. Additionally, the GLA central register could hold analysed data or pattern data, both of which would be non-personal and therefore not subject to GDPR. This could either be one data trust per borough or possibly project specific data trusts, say one for energy London-wide or one for transport London-wide. Whatever the practical form the data trust would take, which would be dependent on the circumstances and its intended future purpose, the legal model could be the same.

It should be noted that whilst the above structures are the suggested forms as to what a data trust could look like, the GLA and RBG did suggest other forms that could meet their data sharing requirements without the need of creating a separate data trust. The first, suggested by the GLA, was to have a central register of data with a standard form of data sharing agreement which parties can use to share data, with each data provider holding its own data and approving each party requesting access. This puts the governance, responsibility, risk and overall motivation on the individual data provider and is only slightly building upon the current system by allowing users to see what data is available. The second model, suggested by the FCC, was to have an individual who is effectively an adviser who can facilitate data sharing and advise on what good data sharing practices are. Again, this is similar to the current system and is greatly dependent on the participation of companies to make initial approaches and manage engagement with each other when it comes to data sharing.

*Article 4(1) of GDPR*
Providing Data to the Trust

The data trust model will only be effective in sharing data if data providers feel confident in supplying their data to the data trust, and so long as they can do so without breaching their legal obligations.

RBG, as the primary point of contact for both the suppliers being used in the energy case, and as the landlord (or freeholder) of the dwellings, will most likely hold individuals’ personal data. For the energy use case, if the data gathered from each of the tenant’s flats, does not contain specific addresses or other such identifiable information about a person, the heating data would not fall under the definition of personal data and therefore would not be captured by the provisions of GDPR. Even this simple example should be treated with some caution. Any data user who had access to billing data would be able to cross refer usage amounts with this to identify individuals and, in addition, anyone with other relevant data sets could potentially carry out a similar exercise, thereby making the data personal again and subject to GDPR. As it currently stands, the energy use case aggregates data at the system-level without the input from a person, therefore an individual tenant should not be personally identifiable. If this is the case, such information should fall outside the scope of GDPR.

Similarly, at this stage with the transport use case, on the condition that no personally identifying data is being shared, the data gathered would also not fall under the provisions of GDPR. The parking sensor data is not likely to be personal data as it just detects whether there is an object in the parking space, regardless of whether it is a car or not, as the sensor does not have this functionality. For the smart electric vehicle charging data, the supplier would likely be unwilling to share payment data or information about individuals and the data trust and prospective data users would likely have no interest in this data. More appropriate and useful will be usage statistics, peak user areas, average use time or popular times of day of use, none of which needs to be personally identifiable. In fact, the kind of data appropriate for sharing will likely form the usual metrics that the companies will use for their own internal analytics (which likely carves out personally identifying information as this is generally irrelevant for the purpose of data analysis).

Personal Data

The GLA has expressed a strong wish that it does not handle personal data. Their particular area of interest is in aggregated data and trends and they would not wish to have the responsibility that would be associated with handling personal data on behalf of a large number of individuals. They do hold some (albeit very little and occasional) personal data, such as a list of school age students, although this data is unrelated to the two use cases being focused on. However, this data, which is used to predict how many schools will be needed over the next few years, is data which the GLA have already agreed not to sell or commercialise in any way.

Whilst there is otherwise no personal data being shared currently and potentially none in the future, RBG, who could be the receivers of personal data before aggregating it and passing it onto the GLA, might need to future-proof for a situation where personal data is being shared. For this they would need informed consent to the sharing of the personal data and specifics given as to the particular circumstances. This is because the consent provision for processing data is likely the best available option due to its engagement of data providers and there being no issues of having to justify to data providers why their data was used without their consent.

27 Article 4(1) of GDPR
If there is no immediate intention to share the personal data it would be difficult to scope an informed consent to its sharing in lines with the provisions under GDPR. Potentially, if there are a few circumstances when personal data would be shared then consent can be given on an ad hoc basis without incurring consent fatigue. This also ensures maximal engagement on the part of individual data providing members of the public.

Under the current system, RBG may gather data for the energy use case under heating regulations which allows such data to be collected in order to enable accurate billing dependent on usage rather than as a flat rate. It should be noted that the legislation only permits RBG to gather temperature data from the flat concerned, it would neither cover the gathering of humidity data nor the sharing of any data. For these, the express consent of the tenants would still be required. Such explicit consent will also be needed if the sensors gather any data that can be used to intuit anything under special categories of personal data which includes racial or ethnic origin, health or biometric data.

**Anonymisation and Pseudonymisation**

As previously mentioned, the interest of the GLA is in aggregated pattern data which circumvents this issue. If RBG uses effective anonymisation or pseudoanonymisation techniques, then any information passed onto the GLA will not be classed as personal data as long as it cannot be used to identify a "natural person." There are limitations to this in that cross referring one person in a data set to another and compiling the results would require each to be individually identifiable. In theory though, this should be fairly limited in scope and potentially can be dealt with on a case-by-case basis. Stripping out any personally identifying information prevents the burden of compliance and therefore the risk of non-compliance being placed on the GLA who would rather that trend and pattern data is shared. Additionally, if a data set is being used to train AI, as some within the GLA and RBG have suggested a data trust might do, data without personally identifying names or locations of people should be sufficient for training data; if it is not, engagement with the particular data provider(s) can be carried out to scope out the reasons why personal data is needed for the particular use case. It is difficult to rely on the anonymisation approach. However, due to there being several de-anonymisation techniques, thus there is the potential for the ICO to still regard this as personal data. The best way to approach this would be by requesting that the ICO issue an opinion on whether satisfactorily anonymised data can fall outside the purview of GDPR, even if it theoretically possible that they can be de-anonymised with advanced techniques.

This means, practically speaking, that RBG, as the data provider of the data from the temperature and humidity sensors relayed on from the supplier, would have the responsibility for complying with GDPR and stripping out any personal data if the data was to be shared more widely. Any consent would therefore only have to cover the collection of personal data for the stated initial purpose for which RBG originally sought to collect it. It would not need to cover all the possible options for the sharing of personal data as only non-identifying or aggregate data would be shared more widely.

This is likely to be similarly so for, the transport use case. This involves commercial suppliers, all of whom would likely be reluctant to share personally identifying customer data, both due to the risk of customer claims for breach of GDPR, but also due to the reputational risk they would face and the public reaction to their personal data being shared, this being a particular concern to the electric car club. With the current scope however, the data shared does not need to be personal in order for the suppliers to derive a benefit from the data being shared. For example, car club members do not need to be individually identified, only whether a car club car is parked in the wrong space as identified by the smart parking sensor.

---

28 Article 4(11) of GDPR
29 The Heat Network (Metering and Billing) Regulations 2014
30 Article 9(1) of GDPR
31 Recital 26 of GDPR
32 Article 4(1) of GDPR
Commercial Confidentiality

Regarding the energy use case, as the receiver of the supplier’s passed on data and provider of the data if it was to be shared, RBG does not have a duty to keep any commercial confidentiality to any third-party commercial organisation. The suppliers in the transport use cases would additionally not have any confidentiality to keep to commercial actors as, as it currently stands, no Intellectual Property Rights, confidential client/business partner information or trade secrets would need to be shared. It would seem that the only suppliers that might have potentially sensitive client information capable of being shared is “Enterprise”, as operators of the electric car club or, for a similar reason, Blue Point London Limited, trading as Source London as suppliers of the electric lamppost charging stations. Any ability to share data will be contained under the contractual agreement under which members use the car club. Without being able to see the content of any such agreement it would be difficult to judge whether the contractual arrangement would cover the sharing of data. If the data is changed however and turned into generic data so as not to be personally linked to any one customer, or into analytical data for Enterprise or Source London’s own use that can then be shared; it is then likely that no duty of commercial confidentiality will exist.

Any data provider that subsequently joins the data trust will have to ensure that any sharing of information with the trust does not breach any commercial confidentiality that it is required to keep. If this is an issue, then there could be an amendment to the data-set so it falls outside the definition of commercially confidential information (i.e. falls outside the protection of Intellectual Property Rights or other protections). This is however difficult to achieve without changing the nature of the data and might produce friction between the commercial organisation and the party to whom it owes a duty of confidentiality. It is likely to be more effective to engage with the party who has the benefit of confidentiality, about sharing their data.

As public bodies, neither RBG nor the GLA have information that can be used for insider trading (although the information they receive from private enterprises can have the potential to be subject to insider trading regulations). There is however the potential for the transport use case suppliers to be captured by insider trading legislation. Information such as how much uptake there has been of the electric car club or the popularity of the electric vehicle charging stations might be used by data users to gain the inside track on the relative successes of parts of the data provider’s business, which investors could use to influence their investment decisions. The risk of falling foul of the insider trading is on the person who is using the information to gain a benefit as long as when the data trust or commercial supplier is disclosing the data, they do so through the proper process and as part of their office, employment or profession. Any data trust rules for accessing data will have to ensure that the data user is clear on what they can and cannot do with that data.

---

33 Article 8 of the Market Abuse Regulation (Regulation 596/2014)
34 Criminal Justice Act 1993
35 S.7 Market Abuse Regulation
Rights in the Data

As it stands, there are no third-party intellectual property rights with the transport and energy use cases that would be shared as part of the data trust. Given the intention of the GLA and RBG to use a data trust in some form, having third party intellectual property rights in the future would seem unlikely as well as they are specifically looking to use data to promote London’s economic growth and use the interconnectedness of data shared to enable better decision making. The GLA in particular, see this as a key resource for informing city planners as London grows.

If such a scenario were to arise in the future where Intellectual Property Rights were being shared within the data trust then there would need to be an assignment or more likely a licence of the Intellectual Property Rights, either being open-ended or to revert when triggered by a future event or after a specific time. As with any sharing of third-party information, the data provider will have to be careful about maintaining commercial confidentialities held to its commercial partners in addition to its duty to any employees, either under their contract of employment or GDPR.

In terms of its own rights, a data trust should be wary of any arising rights in data. If the GLA has an aggregated source of data then database Intellectual Property Rights might arise in that data as a result of the time and effort, but not the cost, invested in transforming the data into a combined database. This would stand apart from the rights of the data providers licensing their data to the trust and the data trust would maintain the rights in the database regardless of if the data provider revokes its licensing of data to the trust. This does not mean that the data trust can keep the revoked data, rather that they have a database right in the shifting database, even as content is constantly changing.

36 British Horseracing Board Ltd and others v William Hill Organisation Ltd (Case C-203-02) [2004] ECR I-10415

37 Freistaat Bayern v Verlag Esterbauer GmbH, Case C-490/14
Receiving Data from the Trust

Careful consideration should also be given to how the data trust manages the data received. This is because, as the GLA and RBG are public organisations, they have emphasised their reticence over having the responsibility of managing the data.

If the contractual model is applied then a recipient of data will be the same as the end data user. In this model it would be at the discretion of the supplier who could access the data rather than that decision being made by a separate body that made up a data trust.

If an organisational model is adopted the central organisation would be the receiver of the data with the subsequent prospective data users who are seeking to access data being data users.

Accessing data

The GLA has expressed a strong wish that it does not handle personal data. Their particular area of interest is in aggregated data and trends and they would not wish to have the responsibility that would be associated with handling personal data on behalf of a large number of individuals. They do hold some (albeit very little and occasional) personal data, such as a list of school age students, although this data is unrelated to the two use cases being focused on. However, this data, which is used to predict how many schools will be needed over the next few years, is data which the GLA have already agreed not to sell or commercialise in any way.

Whilst there is otherwise no personal data being shared currently and potentially none in the future, RBG, who could be the receivers of personal data before aggregating it and passing it onto the GLA, might need to future-proof for a situation where personal data is being shared. For this they would need informed consent to the sharing of the personal data and specifics given as to the particular circumstances. This is because the consent provision for processing data is likely the best available option due to its engagement of data providers and there being no issues of having to justify to data providers why their data was used without their consent.

Restricting Access

Restrictions on who can access data should be less of an issue for the energy use case as there are no commercial holders of data and the GLA and RBG, as public bodies, cannot fall foul of competition law. Rather, it will be the individual data trust holding the data, if an organisational model is followed, that will have to ensure it is compliant with the provisions of the GDPR.

In terms of competition law issues, if the data being shared are usage statistics or other such analytical and meta-data then the nature of the data being shared should not be of the type that could give a market advantage to a competitor. This should both satisfy the commercial supplier that the data being shared will not give a competitor a commercial advantage, but will also mean that the data provider should not fall foul of competition law as a result. If the data being shared were more akin to a trade secret, not only will a commercial supplier be unlikely to share this but, depending on who they share it with, there could be competition law issues if one prospective data user is arbitrarily granted access over another.

Potentially data, such as user numbers of car club members, has the potential to be commercially sensitive data that Enterprise would want to restrict competitors from having access to, however it seems unlikely that to prevent competitors from seeing this would amount to abuse of a market position.

38 S.2 of the Competition Act 1998
39 Chapter I Competition Act 1998
40 Chapter I Competition Act 1998
41 Chapter II Competition Act 1998
Onward Disclosure, Freedom of Information and Subject Access Requests

Equally, again with commercial suppliers and with individuals providing personal data, they would want there to be a restriction on the use and onward disclosure of data to other third parties. This is to protect the data provider to ensure that their data is being used only in such a manner as has been preapproved by them. If data was licensed to the data trust by a data provider, such restrictions on use could be contained within the individual license itself.

Therefore, any breach by any data user accessing the data could be enforced under the terms of the licence. In order for this to be the case, the terms of the license from the data provider to the data trust and also the license from the data trust to the data user, would need clauses that have mirrored provisions allowing for such enforcement. This means that no single form of license would be possible as different datasets would require different licenses. Additionally, prospective data users would have to abide by the terms of use of the data trust. Having both options means there could be general trust provisions which data users must abide by in addition to any further restriction imposed by the data provider under the license terms.

As the data trust will be holding data, potentially specific to individually identified persons then, if an organisational model is followed where data is held by the data trust, individuals will have a right to make requests to determine what data is held by the trust. RBG receives a high volume of Freedom of Information Act requests. Any data trust would likely not be a public body and therefore any request made to it to disclose what data it holds about an individual would be conducted under a subject access request.

Whilst this might seem onerous, this should in theory not be an issue to the data trust as there would have been engagement with the data providers supplying their personal data before their data is provided to the data trust. If the data is anonymised or aggregated, so as to be no longer personally identifiable, then it is not capable of being requested under either of these forms of access. This would mean that, as the GLA would only wish to hold aggregated or analysed data, such requests would not be made to them, rather in all likelihood, to the borough or project specific data trust that operates more locally.

“If data was licensed to the data trust by a data provider, such restrictions on use could be contained within the individual license itself.”

42 https://www.gov.uk/make-a-freedom-of-information-request
**Data Security**

Given that the GLA and RBG are generally at this time not looking to disclose personally identifiable data, or data which would be commercially catastrophic if it was leaked and spread more widely beyond the prospective data users, there should be limited negative impact if data is disclosed to an incorrect third party. One particular use of data that the FCC has suggested is as a data set for training AI. The type of data that would be used for this is unlikely to be commercially or personally sensitive but can be of great value for the type of repetitive data set needed to train an AI through machine learning.

Any breach of the data trust rules, or data licensing agreement between the data provider and the data trust, can all be enforced in the normal manner through the courts or via arbitration. In order to provide full assurance to any data provider and expediency in the case of any breach, then methods of alternative dispute resolution (referred to below) should be considered to resolve disputes, with such provisions being written into the data trust rules and the license agreements.

The GLA and RBG have made very clear that they would like to see assurances from the data users that they have the appropriate organisational structure and security measures to ensure that data will not be leaked to any unauthorised third parties. The onus can be on the third-party body that forms the data trust to ensure that such technical measures are in place before data is shared with a data user. The general impression from the interviewees spoken to is that the GLA and RBG wish to comply with a higher level of data protection standard than the minimum standard suggested by the ICO.

The ICO usually recommends, and can even require, an organisation carrying out “data protection impact assessment” when a new project is started. For a data trust this could be conducted when a new data provider or data user is added to the data trust. Whether the individual data trust wishes to do this will entirely depend on whether RBG, the GLA or the separate body that makes up the data trust think it prudent to conduct this or whether it feels assured that the data is secure by the prospective data user or provider having firm and clear data security policies. The functionality of such methods could, depending on the technical architecture and resources available to the data trust, be built into the platform for accessing the data to ensure security. Ultimately though, as each individual data user of the GLA and RBG data trust will want to carry out analytics on data and use it for their own purposes, they would need to transfer the data from the data trust platform to their own, for which they would need robust technical structures in place to ensure security.

---

46 Article 35(3)(c) of GDPR
47 Article 35 of GDPR
Meta and Derived Data Ownership

One key point that is very difficult to resolve is who owns the rights in the derived or meta-data, and any research results that arise out of analysis of that data. In one sense this will be decided by what is in the license agreements and a matter of negotiation between parties when access is proposed, if this is done on a case by case basis then this is possible to agree who has ownership of the meta-data. Regarding the energy use case there is a "systems energy management system" that is being developed by Imperial College who we believe will have access to some of the energy use data in order to trial the system and analyse some of the data.

If the data trust aims to create a universal trust standard for all its data providers and users, then either the product of any data analytics belongs to the data user conducting the analysis, either the GLA or a third party, or the data provider retains an interest in the results of the data analysis. For the purposes that the GLA and RBG are suggesting for the data trust, they would potentially wish to see some retention on the part of data trust of any data analytics. If the data used for the analysis is paid for, then a monetary benefit could be returned to the trust whilst if the data were to be used for a more prosocial benefit, then it could be more freely available.

It would seem likely that commercial suppliers, or anyone paying for access to the data, however, would wish to retain the results of any data analytics, unless a third party were willing to pay them for access to these results. Whichever the preferred method is for the data trust will be entirely down to what the stakeholders wish the structure to be, but either option is viable. From a pure intellectual property law perspective, the rights in the products of the analysis will be in the creator of the intellectual property, although this can be overwritten by any agreement that supersedes this and details any intellectual property provisions. In a traditional context, this is usually in the form of a service agreement to conduct research on behalf of the data provider or an employment agreement with provisions to retain any intellectual property. For each of these traditional contexts however, the data provider is paying the data user to produce the intellectual property, and in the data trust instance, the data user is paying for or otherwise gaining access to data and producing the intellectual property for its own benefit.

In terms of the transport use case, the commercial suppliers will likely be uninterested in the result of any data analytics as they would more likely than not being paid for the data. Conversely RBG and the GLA might be more interested in the results of any analysis that can be used by them and potentially shared more widely for a possible prosocial purpose. Whether this would be possible is whether the GLA, RBG or the data trust, retain any rights in the intellectual property being produced, otherwise they would have to obtain the consent of the analyser of the data, and therefore the owner of the intellectual property, to share. Ultimately this would be a point of commercial negotiation between the parties and it would seem difficult to have a universally applicable approach as each party will likely have a preferred option.
Re-providing Data back to the trust

Given that the GLA and RBG have a preference that any data trust they are involved with is one which has a primarily prosocial purpose, there could likely be scenarios where altruistic data users wish to donate data back to the trust, or even that donating back is a requirement of being granted access. As established in the general legal report, data is not a physical asset capable of being donated. As such if the licensing model were to be adapted to account for the fact that someone wishes to donate their data to the trust, the license could take the form of an irrevocable licence to use the data indefinitely and retain any product of analysis producing intellectual property. Such a non-exclusive license would be similar in nature to the type of open-source licensing format that exists currently for software development. Given the prosocial nature of the GLA and RBG, this scenario is potentially conceivable if people feel assured the trust is working towards prosocial and community minded goals.

It should be noted that in the case of the transport and energy use cases these are unlikely to benefit from data being donated to the data trust s concerned. Both use cases benefit from live data being shared in real-time to build up a picture of patterns and trends. There is less of an onus on analysing historical data. More relevant to the concept of donating data would be for research organisations or other such potential providers who have large data sets that can benefit from analysis and cross-referencing with other data, a concept which is more relevant to the wildlife data trust pilot. As data is not considered an asset then the data trust will not have to deal with any issues relating to gifting assets or tax issue.

https://opensource.org/docs/osd
Ensuring Compliance with the Trust Rules

As publicly minded and funded organisations, the GLA and RBG are particularly aware of how data can be misused by bad actors and how it is very easy for the public to be wary of new organisational structures that hold data.

The general public, which any data trust involving RBG and the GLA will only be one step removed from, are particularly sensitive to misuses of data by commercial and public organisations and thus are likely to be reluctant to engage with a structure that facilitates the free sharing of data.

As a result, a commonly echoed sentiment from both bodies is that the standards by which a data trust should operate should be unimpeachable and higher than whatever is the legal minimum standard. Part of this is ensuring strict compliance by the data trust, data providers and data users to whatever is the agreed upon standard terms of use.

https://www.theguardian.com/uk-news/cambridge-analytica

https://www.theregister.co.uk/2018/12/05/nao_windrush_liberty_data_sharing/
Alternative Dispute Resolution

Keeping this in mind, the data trust will likely wish to embrace a form of Alternative Dispute Resolution ('ADR'). All of the potential methods of enforcement referred to in the general legal report, such as contractual claims for breach of licence terms or breach of GDPR, against the data trust, are still available for the use cases within the RBG and GLA pilot. Additionally, contractual claims against bad actors who misuse data they have accessed, either directly or indirectly, or between data users, can still be pursued. As referred to in the general legal report, these methods though are expensive, time-consuming and, by their nature, adversarial. This would not necessarily be conducive to resolving a dispute with a view to have a continued data sharing relationship.

Conversely, using ADR as a method for resolving a dispute could prove to be expedient; it is cheaper, more collaborative and in line with the GLA and RBG’s wish to have a self-governing and cooperative organisation.

Of the ADR methods suggested a few could be potentially applicable to the two use cases. Arbitration, adjudication and expert determination would likely be unsuitable for RBG and the GLA for the same reasons that they would be ill-suited to a data trust in general, the reasons being referred to the in the general legal report. Of those options that were deemed less optimal in the general legal report because they are non-binding, some could be applied to the RBG or GLA data trust due to their willingness to abide to a high ethical standard and the potential public pressure if they do not do so. Mediation or conciliation could be applicable to the energy use case as there are no commercial actors (yet) involved and there must be a high degree of stewardship and engagement. There is also a greater need, seeing as the data providers are tenants and the proposed manager of the data trust their landlord, to have a higher level of engagement. Mediation/conciliation would mean that there would be a mutually agreed upon penalty accepted by both parties for any breach of the trust rules. If this is not agreed with or a resolution cannot be reached, the matter could be referred to a more determinative method of Alternative Dispute Resolution of the type that would be more suitable for the transport use case.

The ADR method suggested in the general legal report as being most suitable for data trust’s generally, the Dispute Resolution Board, will likely also be the most appropriate option for the transport use case. This is because if there is a breach of the trust rules by one of the commercial suppliers, they are more likely to either dispute fault entirely, or only accept the non-binding suggestion if it is skewed in their favour. Therefore, having an option that makes a determination that is contractually binding, but that can be challenged through the courts, would be better suited when there are the conflicting commercial and prosocial interests of business enterprises and public bodies.

For either mediation, conciliation or a Dispute Resolution Board, independent and neutral third-party individuals will be needed to facilitate a resolution. The cost of these could be funded through money generated from any commercialisation of the sharing of any data.

Having these internal mechanisms for resolution are important as RBG and the GLA have made it very clear that they do not want to be responsible for any oversight of these data trust’s beyond keeping to the agreed upon standard set of trust rules as a stakeholder in the trust. As public bodies themselves, whilst they are not against some sort of auditing function by another government body ensuring compliance and best practice, there is as well a general sentiment that the data trust will be more adaptable if compliance is internal, and they will be viewed as generally more ethical and trustworthy if they are able to positively govern themselves to a high standard without enforcement from an overseeing government body. It should be noted that in any case they will have to abide by any audit conducted by the ICO, although this will not ensure compliance to trust rules, only the minimum legal standard relating to the managing of data.

a) https://ico.org.uk/for-organisations/resources-and-support/audits/
Certification and other forms of Compliance

One particular method of ensuring compliance that was not suggested in the general legal report, but has arisen during the user interviews as something that the GLA and RBG might have an appetite for, is certification of good practice for data trusts. This is relevant to compliance with trust rules; not only is it a way to ensure good practice on the part of the data trust but it also incentivises the data trust to rectify any breaches in order to maintain their status as an officially certified data trust, and can be used on the part of the certificating organisation to ensure that the data trust is complying with good data standards. Evidently, such certification would have to be undertaken by an independent third-party organisation with an expertise in good data sharing practices, of which the Open Data Institute is one. It is recommended that further consideration be given to the development of such a certification process.

The final potential method of compliance, particularly relevant to the transport use case commercial suppliers, is that if the data trust is acting in a way with which a data provider disagrees, the commercial supplier can withdraw its data from the data trust under the provision of the agreement under which it licensed its data to the data trust in the first place. Whilst this does not provide rectification or compensation where there has been a breach of the trust rules, the threat of a data provider pulling its data from the trust could practically ensure the data trust’s compliance with ethical standards of data sharing and the trust rules.

The GLA have suggested for the energy use case that if the tenants receive the benefit of the heating sensor, which are a potential reduction in the amount of energy used due to smart heating analytics, then they should also consent to their data being collected. As referred to previously, such data can be currently obtained through the heating regulations to ensure accurate billing and thus is covered by the provision of GDPR that allows the collection of data to meet the controller’s legal obligations. This however would purely cover the collection and not the sharing of personal data. Sharing of personal data beyond the data collector, in this case RBG, is unlikely to fall under any of the six lawful reasons under which an organisation can process data save for the consent requirement. For there to be any sharing of data deemed to be personal therefore, consent of the data providers, in this case being the residents, would be needed. There is an argument that processing could be granted under the legitimate interest of the controller or public interest justifications within GDPR; however, these have tended towards a narrow interpretation and it is not clear whether these could apply in a data trust context.

---

52 Article 6(1)(c) and Recital 46 of GDPR
53 Article 6(1), Article 6(2) and Recital 40 of GDPR
54 Article 6(1)(f) of GDPR
55 Article 6(1)(e) of GDPR
Governance Structures and Operations

Once a data trust has been set up and the data has been received and is being held and secured correctly, the data trust will have to have methods of governance that are effective to ensure its goals are met and to balance the interests of any stakeholders.

As referred to previously, from the interviewees spoken to it is generally considered that the GLA and RBG would prefer a data trust to be as self-governing as possible, both as a way to ensure it is viable and self-sustaining but also to show it does not need a third-party overseeing organisation enforcing compliance. In the case of the CIC model, the regulator would make sure that the data trust is pursuing its prosocial purpose, but it would not dictate the governance structure of the CIC or have any involvement in enforcing other rules of the data trust.

Board and Partners

In a corporate body or partnership, there is a governing body in the form of boards of directors or the partners themselves, who make everyday decisions on behalf of the organisation. If stakeholders are members or shareholders of these organisational structures then the directors or partners will have to run the organisation for the benefit of them and they will be able to vote on more major, but less frequent, matters. A contractual model does not automatically have this type of governing body created around it, rather any management of the agreements would be between the parties and either governed by the terms of any data sharing agreement or negotiated between stakeholders.

The view of the GLA and RBG appears to be that the data providers would be those with a say in the running of the data trust as it will be their data that the trust handles. Potentially the data providers would be the shareholders or members if the corporate model was followed, thereby giving some input in how the organisational structure is managed. This would suit the transport use case as there are a few large commercial organisations and the public bodies of the GLA and RBG, all of whom are providing similar levels of contribution and therefore each could have representatives who could sit on the board of directors. The energy use case might not so easily translate to a board governance structure unless there were representatives of the tenants, perhaps from tenant associations, who could sit as non-executive directors on the board. It would seem likely that the data users would not have a say in the governance of the data trust, although this does not have to be the case, rather they would apply to the data trust for access to the shared data.
Achieving Legitimacy

Legitimacy of the data trust organisation will only be possible if there are strict governance structures in place, as the GLA and RBG wish to hold, and be seen to be holding, themselves to a high standard of trustworthiness. The commercial suppliers will need to feel assured that any data shared is managed correctly and that they will have some say in the way in which it is shared. Commercial organisations will be concerned to know that brand reputation is maintained and contractual or legislative breaches are avoided by the commercial organisations when they are licensing their data to the trust.

Legitimacy in the mind of the public, which is of particular relevance to the energy use case, will be that when data is being shared it is not to be used for illicit means, targeting advertising, being shared unjustly with the police or immigration, a particular concern of the policy strategist for the GLA, or that energy use data can be used by bad actors to determine when someone is at home and therefore can determine the most opportune moment to burgle a residence. In order to achieve such legitimacy, in all likelihood the data trust will require some level of transparency in the way it operates. Information held by the data trust would have to be disclosed to the relevant police or immigration authorities on proper application to the courts. However, this power to demand production of information exists regardless and the police could make the request to the data provider in any case.

Certification or a clear agreed upon standard for the data trust rules will be an effective way to ensure potential data providers are convinced about the legitimacy of the data trust organisation. There have been various suggestions of the level to which a data trust should operate, whether it is a London-wide data trust, borough trust specific to RBG, an electric vehicle, social housing, smart city or even an Open Data Institute managed data trust. Regardless of the level at which the data trust(s) will operate, the governance mechanisms inherent to corporate organisational structures will be applicable. Given the public nature of the organisations and the fact that the government is generally supportive of the principle of ethical data sharing, their reputation is likely to rub-off on the data trust and increase its legitimacy in the mind of the data providers and data subjects. This would be more so than if the data trust were, for example, set up and managed by a group of commercial actors, whom the public might view with greater suspicion.

External representation

Although self-governing data trusts provide one form of governance, often with institutions in quasi-public sector organisations like the care sector or the prison system, external regulators are used to ensure strict compliance to industry rules to facilitate good practice.

Whilst a data trust need not be a publically influenced body, as the data trust for this pilot will either be set up by the GLA and RBG, or at the very least influenced by them, then some sort of external overseeing organisation would likely be more in line with what government bodies are used to and are likely to inspire trust in the mind of the general public.

Paragraph 1 of Schedule 1 of the Police and Criminal Evidence Act 1984
Certification

If external governance is used, more likely ensuring compliance above a minimum legal standard of how an organisation generally should handle data, this would come from an external third-party organisation that can provide certification on what good data practice standards are and which can certify whether or not the data trust is following them. The idea would be that the implementing organisation could create a certification standard similar to what the British Standards Institution does for companies more widely.67 This would not only mean that a data trust does not have to wait for government mobilisation to set up standards of good practice but any payment would come through the form of a certification fee to the third-party organisation.

There are some forms of certification that cover good data protection standards both issued by Bureau Veritas and British Standards,68 however these relate to general data protection standards and would not cover the specific issues that would relate to a data trust. Whether these organisations or an organisation with more intimate knowledge of data trusts would take up the mantle of certification would be a case of seeing which, if any, were inclined to do so. It is outside the scope of this report to provide a detailed analysis and recommendation of a preferred certification vehicle. In any event though, certification will not only ensure that good ethical standards are followed, but the threat of removal of the certification standard in the event of any breach (see above) will help to ensure that any error on the part of the data trust is handled correctly. Individual data providers could also ensure that good standards are followed by the data trust through using the ultimate backstop of having the right contained within the licence of their data to the data trust, to remove that data if good data sharing standards are not complied with.

“The idea would be that the implementing organisation could create a certification standard similar to what the British Standards Institution does for companies more widely.”

69 https://whatis.techtarget.com/definition/BS-100122009-British-Standard-100122009
Corporate Governance and Compliance

It should be noted that a CIC is already regulated to ensure that it is meeting the community purpose that it stated upon applying to the Office of the Regulator of Community Interest Companies. Whilst this does not cover the operational specifics of what would be involved with running a data trust, it would ensure that the CIC data trust complies with the general golden thread of promoting data sharing in an ethical manner.

There are no current ombudsman services that specifically relate to data; breaches of data protection would fall generally within the purview of the ICO. If the form of the data trust was a body that was publicly run, any breaches of its duties or stated aims would be captured by the local government ombudsman. If not, and it were a private body then the data trust could be captured under the scope of the Trading Standards Institute or another such regulator or public body. As previously mentioned, these would ensure good business practices generally but are not specific to data trusts. It is recommended that thorough consideration of the various ombudsman type services available be undertaken prior to any decision being made in this regard.

It should also be noted that if a corporate model is followed for the structure of a data trust then there are requirements to publish company accounts at Companies House. The particular structure chosen will determine the level of disclosure within the accounts. However, any accounting information published will ensure a measure of public scrutiny, again providing a reputational pressure to comply with a good ethical standard.

Additionally, a particular standard of practice can be inbuilt into any organisational structure's governance documents, be they the memorandum and articles of association for a company, the partnership agreement for an LLP or the club rules for an unincorporated association. This would ensure not only that the organisational body that manages the data-to-day operation of the data trust would have a particular level of governance in mind, but also that the shareholders, partners or club members have a method of redress if the governance documents are not followed.

---

60 https://www.gov.uk/data-protection/make-a-complaint
61 https://www.lgo.org.uk
62 https://www.tradingstandards.uk
63 Companies Act 2006
64 Limited Liability Partnerships Act 2000
65 Blackpool Marton Rotary Club v Martin 62 TC 686
Termination/Winding Up

It is clearly the GLA and RBG’s intention that, once the data trust is set up, it will continue to run indefinitely. As the project is funded by the Sharing Cities programme, with the potential that this funding could be withdrawn post-Brexit, discussion of what would happen in the event that the data trust would dissolve is important. Whilst the structural design of the data trust is that it is self-sustaining, it might be that data providers’ enthusiasm would wane if RBG and the GLA removed themselves from being involved with the data trust.

As the suggested corporate organisational form of a data trust is, in itself, its own body, data providers could opt to remove themselves from the data trust without needing to dissolve the whole structure. This would likely be through a clause contained within the agreement licensing their data to the trust, to allow them to revoke this license and withdraw their data from the trust at any time. It should be noted that whilst revocation of agreements under the contractual model is also possible, this would require any body holding data to remove it from their system rather than from just from the data trust’s own system. For both, there would need to be terms that ensure that information is deleted from the system of any organisation that has accessed data and is conducting analysis on their own systems of the data.

If an organisation did not do so it would be in breach of the agreement, and the rights under it could be enforced accordingly. Any new IP (e.g. those protected by database rights) produced by analysis of the data would likely be the data users to retain. This is because the GLA and RBG are seeking to promote London’s economic development, and they could not do so if a party had gone to the effort and expense to conduct analysis on the data only to find that the data provider wished this to be deleted as they no longer wished to be part of the data trust. Ultimately though, any relevant permutation can be contained within the licence or contractual arrangement between the stakeholders.
Terminating an Organisational Structure

In the event that the structure is an organisational one, then there are pre-established avenues by which companies can be wound up or partnerships are dissolved. Given the intended purpose of the data trust, there is unlikely to be an issue with having numerous creditors demanding payment in the event of insolvency as the organisational structure is not a trading body but rather a form around which the functioning of a data trust can develop. In any event, the liability of the stakeholders will likely be limited in the event of insolvency either due to the structure being an LC, CIC or LLP. Also, given the prosocial purpose of a data trust, provisions can be put into the data trust governance documents that there are no distributions of trust assets to shareholders upon dissolution of the data trust; rather, that these are either donated to another data trust or some other prosocial organisation. This is particularly in line with the GLA and RBG's hopes of a data trust as they do not view it as an inherently profit-making institution and they think if it is, this would mean that the public could view it with distrust.

If the organisational model opted for is a CIC, there is a limitation under the regulations that, upon insolvency, no distribution can be made to the stakeholders of the CIC beyond the original paid-up value of their shares, typically a £1.00 nominal value. If the articles do not specify a beneficiary to whom any residual monies should be left upon dissolution it will be for the Office of the Regulator of Community Interest Companies to decide where the subsequent funds should be distributed, usually to a similar like-minded organisation.

Ordinarily, in the event of insolvency, any assets owned by the organisational structure will be sold and distributed to its members. Where the data is licensed to the data trust from a data provider however, in the event of insolvency there will likely be a provision in the license that ensures that the license ends and the data is returned to the data provider. This prevents the rights under the license, and therefore the data, being sold on to unapproved third parties in the event of an insolvency event. This is in line with RBG and the GLA's stringent requirements that anyone providing data to the trust will retain as much control as possible and not have their data sold on without the data providers approval. Such a provision would be the same regardless of whether the insolvency event was voluntary or involuntary.

RBG and the GLA are likely to be risk averse when it comes to matters that could potentially cost the tax payer money. As a result, they would understandably want assurances that no liability would fall on them in the event of insolvency. As discussed above, if they were stakeholders in the organisations, it would be likely that any model, organisational or otherwise, would contain provisions that limit their liability. If representatives of the GLA and RBG were to have a say in the governing of the data trust they would also want assurances that in this role little or no liability would attach to them (as long as their actions are appropriate). If a director acts within the powers granted to him by the company articles, contractual data sharing arrangement or in good faith to the provisions under the Companies Act 2006, there would be no personal liability to the nominated representative for breach of their legal duties. Thus, it is eminently achievable for RBG and the GLA to have appropriate representation without an unacceptable degree of risk attaching to their representatives.

If, rather than the complete winding-up of the data trust occurs, individual data providers "drop out", say for example one of the suppliers in the energy use case, if it is practically feasible to continue the data trust, any of the models allow the form of data trust to continue with a reduced numbers of data providers, again as long as the minimum operational costs are met to keep the organisation operational. This is particularly useful with the transport use case, as there are multiple commercial suppliers involved who may wish to cease to be part of the data trust, for example if they no longer wish to continue supplying their product in London or to the RBG.

---

66 S122(1) of the Insolvency Act 1986
67 Partnership Act 1890
68 Regulation 23 of Community Interest Company Regulations 2005/1788
69 Regulation 23 of Community Interest Company Regulations 2005/1788
Summary suggested model

The above section of the report represents the analysis of the data trust form from the general legal report, as applied to the form of data trust which, from our various stakeholder engagements, the GLA and RBG seem to wish to have established. The following is a brief summary of the proposed format that would meet the disparate needs which both bodies are looking to derive from a data trust.

Potentially Recommended Model

The best suggested legal structure would be a CIC whose stated purpose is to promote the fair and ethical sharing of data. This is so that there would be a body with a separate legal personality that could hold data, yet that has an overriding objective to promote the fair and ethical sharing of data enforced by the Office of the Regulator of Community Interest Companies. Data would be licensed to the data trust from data providers, for which they could receive a fee when it is accessed, a proportion of which is used to fund the data trust, with more prosocially minded organisations either being charged a reduced fee or receiving free access.

A LC has been discounted because, although this structure would be better suited to a group seeking commercial benefit from the data, the impression from the GLA and RBG is that they would like an organisational structure that supports some prosocial benefit to be derived from the provided data. Any disputes would be resolved by a Dispute Resolution Board in the first instance, with the option of the matter proceeding to court if the aggrieved party does not accept the result. Good data standards would be agreed initially but could be maintained through certification from a third-party body, or else relying on the Office of the Regulator of Community Interest Companies to ensure community minded practices are followed.

As a CIC, there would be a board of directors made up of key stakeholder representatives who would handle day-to-day decision making with less frequent stakeholder meetings to discuss more important issues. Data providers could give guidelines as to who can access their data, the process of which could be automated. If certain prospective data providers were wary about sharing with the trust, there can be direct engagement either with the data provider or with the board of the data trust. As a CIC, the data trust could operate with few members. However, if the data trust is wound up, the format of the CIC provides that any assets of the data trust, of which there are likely to be few due to the licensing model, are used only for the support of a like-minded or other prosocial organisation.

From the view of RBG and the GLA, RBG would likely run a borough specific data trust, which actually holds the data. The GLA though would manage an index/register as to what data is held within this borough specific data trust. Potential data users would utilise and find the data that RBG holds, as the GLA intend this model to be implementable across all 33 boroughs.
Data Trust

1. CIC
2. License
3. Certified Best Practice
4. Dispute Resolution Board
5. Board and Regulator
6. Asset-Lock on Distribution
Conclusion

Whilst the GLA and RBG are both keen to reap the benefits which a data trust would provide, their view as to the format it should take represents a panacea of issues surrounding data and data sharing. In reality, a data trust, as implemented by the model suggested above, is a framework for the ethical and fair sharing of data between data providers and prospective data users. It should be noted that the suggested model is flexible and the relevant governance documents for the organisational body of the data trust and the licenses providing data to the trust, can be amended.

This would both help to incorporate the deliberative decision making approach recommended in the pilot specific decision making report and ensure that the legal structure continues to align with the generally evolving idea of the new concept of data trusts.

Whilst the initial funding for the use case pilots have come from the Sharing Cities program, the suggested model has the potential to be self-sustaining, dependent on further testing and market scoping, without further funding needed or being propped up from local or central government.

This meets the requirement for there to be a structure that will outlive the political will to promote data sharing when other issues go to the forefront of the RBG and the GLA’s respective agenda. Generally, encouraging legislative reform or engaging the public in the oversight of anything new can be tricky, especially since the body that constitutes the GLA might change completely, depending on the outcome next election and the consequent, inevitable, change in the timetable of priorities.

Accordingly, a third party body capable of being sustained with or without the input of RBG and the GLA is imperative, one that can primarily be self-governing, able to resolve its own decisions and whose sum ideal level of contribution from an external organisation is through certification by a body that has the willingness to maintain involvement in the data trust long after the political landscape shifts. It is because the GLA and RBG are, for good reason, risk averse and conscious of gratuitous spending of taxpayer income that they are keen to have a model that is self-operating with minimal risk and responsibility of their part. They were happy to operate the pilot to see if there were tangible benefits to having a data trust, given that there is a current political appetite to promote this type of arrangement, both from the view of the Sharing Cities funding but also the general excitement about the potential of big data for promoting technology such as AI.

The aim of the above model is to create a form of data trust that is immediately implementable without having to wait for reforms of current equitable trust law or an overseeing public body to provide the regulation. It also seeks to meet the GLA and RBG’s requirements of having a data trust that is transparent, secure and trustworthy on the part of the data users who will be offering up their valuable and sensitive data to the data trust. If executed properly, the suggested data trust should be self-sustaining and hold itself to a higher ethical standard than the legal minimum, whilst still providing a financial incentive, beyond just the benefits of any data analytics, for data providers to contribute their data to the data trust. The suggested model is also applicable to both use cases as it can support having commercial suppliers and ordinary tenants as stakeholders in the data trust.

The above model should achieve this and, if the GLA and RBG are willing, it can be adapted in order to fulfil any further needs that they have around data sharing. Whilst the transport and energy use cases have been the current topics focussed on, many of the interviews with representatives of the GLA and RBG that we have held envisage the potential for a data trust beyond this scope. Suggested areas include for training AI, use by town planners to make more informed decisions about the development of our cities as they grow and even as a way to help manage the, ever increasing, demands on councils to handle adult social care.

All of these and more are achievable with the suggested model of a data trust, with the benefit that it can be adapted to whatever format the GLA and RBG wish it to take.
Let’s talk

In these pages, we hope we’ve given you a flavour of how we can help you and your business, regardless of company size or life-stage. To find out how we can support you, please get in touch and let’s discuss what would work best to help you achieve your goals.

Rob Bryan
Partner & Head of Science & Technology
01242 248228
rob.bryan@bpe.co.uk

Emily Barwell
Solicitor
01242 248487
emily.barwell@bpe.co.uk

Rupert Parker
Trainee
01242 248222
rupert.parker@bpe.co.uk

This project was commissioned and run in collaboration with the Open Data Institute as part of a project funded by the UK Government’s Office for Artificial Intelligence and Innovate UK. It builds on research from the ODI’s Innovation programme funded by Innovate UK. The views in this report are those of the authors.